The aims of this study are to synthesize and produce a single-chain antibody (scFv) of the anti-cytokeratin 8 monoclonal IgG antibody TS1 and to functionally map amino acid residues important for the interaction with its antigen and the anti-idiotypic antibody TS1. The TS1 antibody has been shown to be effective in binding cytokeratin 8 (CK8) expressed in tumors in vivo and is proposed to be useful in immunotargeting and/or immunotherapy. The anti-idiotypic antibody TS1 can be used to regulate the tumor:non-tumor ratio. Mutagenesis of certain amino acid residues can be used to alter the affinity to improve the tumor:non-tumor ratio further.

In the present study, the TS1 IgG was chemically modified to specify groups of residues important for interaction with both CK8 and TS1. If important residues were found in the CDRs, they were mutated in the TS1 scFv construct and the effect was studied using ELISA.

The main conclusions drawn from this study are that the important amino acid residues in TS1 for the interaction with both CK8 and TS1 are mainly tyrosines, charged residues and a tryptophan. A central interacting interface was identified with the somewhat unusual participation of residues in the CDR 2 of the light chain. Mutations which resulted in increased affinity to both CK8 and TS1 were also identified.

Karlstad University Studies
ISSN 1403-8099
ISBN 91-85335-43-6
Patrik Holm

Single-chain antibody construction and functional mapping of the monoclonal antibody TS1

Its interaction with the antigen and the anti-idiotype
Patrik Holm. *Single-chain antibody construction and functional mapping of the monoclonal antibody TS1 – Its interaction with the antigen and the anti-idiotyp*
Abstract

Antibodies are proteins with the ability to bind antigens rapidly and specifically. Antibodies consist of several different parts. The complementary determining regions (CDR) are the parts that recognize the antigen are the focus of this study.

The aims of this study were to synthesize and produce a single-chain antibody (scFv) of the anti-cytokeratin 8 monoclonal IgG antibody TS1 and to functionally map amino acid residues important for the interaction with its antigen and the anti-idiotypic antibody αTS1.

Cytokeratins are present in significant amount in the necrotic areas of carcinomas and are not released into the circulation since they have low solubility.

The TS1 antibody has been shown to be effective in binding cytokeratin 8 (CK8) exposed in tumors in vivo and is proposed to be useful in immunotargeting and/or immunotherapy. The anti-idiotypic antibody αTS1 can be used to regulate the tumor:non-tumor ratio. Mutagenesis of certain amino acid residues can be used to alter the affinity to improve the tumor:non-tumor ratio further.

In the present study, the TS1 IgG was chemically modified to specify groups of amino acid residues important for the interaction with both CK8 and αTS1. If important residues were found in the CDRs, they were mutated in the TS1 scFv construct and the effect was studied using ELISA.

The main conclusions drawn from this study are that the important amino acid residues in TS1 for the interaction with both CK8 and αTS1 are mainly tyrosines, charged residues and a tryptophan. A central interacting interface was identified with the somewhat unusual participation of residues in the CDR 2 of the light chain. Mutations which resulted in increased affinity to both CK8 and αTS1 were also identified.
List of original papers


II. Holm P. Jafari R. Erlandsson A. Stigbrand T and Sundström B.E. Functional mapping and single chain construction of the anti-cytokeratin 8 monoclonal antibody TS1. Manuscript
**Abbreviations**

γTS1  anti-TS1  
CDR  Complementary determining regions  
CK  Cytokeratin  
DEPC  diethylpyrocarbonate  
EDC  1-ethyl-3-(3-dimethylaminopropyl)carbodiimide  
ELISA  Enzyme-linked immuno sorbent assay  
Fab  Fragment antigen binding  
Fc  Fragment crystallizable  
FDNB  1-fluoro-2,4-dinitrobenzene  
FR  Frame/non-CDR regions in variable domains  
IgG  Immunoglobulin G  
NMR  Nuclear magnetic resonance  
PCR  Polymerase chain reaction  
PDB  Protein data bank  
PEG  Poly ethylene glycol  
pHPG  para-hydroxyphenylglyoxal  
scFv  single chain antibody  
SDR  Structurally determining residue  
SPR  Surface plasmon resonance  
SDS  Sodium dodecyl sulphate

Amino acid residues are referred to by their one-letter abbreviation.
Table of contents

Antibodies 9
Antibody structure 9
Antibodies in the clinic 11
Modified antibodies 11
Anti-idiotypic antibodies 12

Cytokeratin 12
Cytokeratin in carcinomas 13

Protein interactions 14
Antibody-antigen affinity 17
Tools for studying antibody interactions 18
Antibody modeling 18

Aims of this study 20

Results and discussion 21

Future perspective 26

Acknowledgements 27

References 28
**Antibodies**

Antibodies or immunoglobulins (Ig) are proteins with the ability to bind antigens that include a wide variety of molecules from small carbohydrates to large proteins. Antibodies are an important part of the immune system and protect us from different infectious agents. This means that the antibodies have to discriminate between molecules from our own body (self) and molecules from the infectious agents (non-self).

**Antibody structure**

The antibodies are “Y”-shaped proteins comprised of two identical heavy chains and two identical light chains. All chains consist of a variable region and a constant region, depicted in figure 1a. The heavy chain has four or five domains, depending on antibody class and the light chain has two domains. Each domain is built of two antiparallel β-sheets. The variable region consists of the N-terminal domain from each chain. The constant region consists of the remaining domains. The domains are interacting by hydrophobic interactions and disulfide bonds. The variable domain of one of the light chains is held together with one of the variable domains from the heavy chain and the light constant domain is packed together with the first heavy constant domain. The remaining heavy constant domains are packed together and constitute the Fc-part of the antibody.

The tight packing described above is formed in a way that allows the loops that connect the β-strands in the β-sheets in the variable regions to be positioned close to each other. These loops are referred to as the complementary determining regions (CDR) and are responsible for the antigen recognition and binding [1]. There are six CDRs, three in the light chain and three in the heavy chain.

The CDRs are also called hypervariable regions because the amino acid sequence is very variable. However, there are highly conserved amino acid residues in the CDRs, called key residues or structurally determining residues (SDR), and these residues are responsible for the structure of the CDRs. Theoretically, the CDRs can adopt a wide range of conformations but only a few of these conformations are seen in antibodies. It seems that these conformations have been conserved, through the conservation of the SDRs, throughout the evolution and the antigen binding capacity is lost if any of these conformations not are achieved. The conserved conformations seen in antibodies can be divided into classes that are closely structurally related and are
called canonical classes [2]. The canonical classes are primarily based on the length of the CDR but also on the SDRs. The SDRs are involved in the folding of the proteins and are responsible for getting the CDRs in the right conformation. The SDRs divide CDRs into clusters based on what specific residue a specific SDR consists of [3]. These classes and clusters can be used when the structure of an antibody paratope is to be predicted and modeled.

Figure 1. a: A schematic image of the domains of an IgG. b: Schematic images of variants of modified antibodies. H = heavy chain, L = light chain, V = variable domain, C = constant domain for all images.
Antibodies in the clinic

Antibodies can be designed to be used in a number of ways in the laboratory or in health care. All of the applications are using the antibody’s ability to rapidly and effectively bind to antigens. Antibodies are used in different assays in laboratories worldwide, for example to determine which allergens an allergy patient is allergic to or to measure different hormone levels such as insulin or estrogen.

Another application for antibodies is the use in the treatment of diseases, for example the drugs Herceptin [4] and Rituxan [5] that are both used for tumor treatment. Most of the tumor treatment antibodies are used to treat hematological tumors, but present phase III trials are including several antibodies to be utilized in non hematological tumors for example ovarian and colorectal cancers [6]. This new therapy named immunotherapy has advantages over traditional therapy. It has fewer serious side effects than traditional cytotoxic drugs. The immunotherapy is, just as the traditional treatments, often combined with other treatments such as cytotoxic drugs and radiation.

Modified antibodies

Antibodies can be modified to several different forms, all with maintained antigen specificity. The easiest way is by enzymatic cleavage of the heavy chains which will result in a free Fc-part or partial Fc-part and Fab/F(ab')2 fragments, where the Fab/F(ab')2 fragments have the antigen specificity (figure 1b).

Another fragment is the single-chain antibody (scFv) which is not formed by the proteolytic manner described above (figure 1b). The scFv is one of the smallest antigen binding fragments and is formed by joining a variable domain from each chain to each other by a linker peptide [7]. The linker has to be flexible and soluble and not induce immunologic response. The most common linker used is the (Gly4Ser)3 where the glycine gives flexibility and the serine gives stability and solubility [8]. Other linkers are used but the idea of flexibility and solubility is always preserved.

The amino group in the N-terminal residue of the antibody chain that is in the peptide bond to the linker is sometimes important for binding to the antigen. The scFv construct is then not able to function properly. This problem could be handled in two ways, separately or together. The first way is to change the order of the variable domains which only will work if just one of the N-terminal amino groups is important. The second way is to, C-terminally in the linker,
incorporate an amino acid residue whose side-chain contains a positive charge, for example lysine, thereby compensating for the lost positive charge of the N-terminal amino group. If the first way is considered it is likely that the linker has to be altered. The most common order of the variable domains in a scFv is heavy chain, linker, light chain. For this the 15 residues of the (Gly,Ser)_3 linker is enough. If the opposite order is used, the linker has to be extended to achieve a properly folded scFv.

**Anti-idiotypic antibodies**

Anti-idiotypic antibodies bind to the paratope of another antibody, the idiotype, and hinder the idiotype to bind to its antigen. A true anti-idiotype will, in an immunized animal, give rise to an anti anti-idiotype that has the same antigen specificity as the idiotype. If this does not happen, the anti-idiotype is a false anti-idiotype and just partially overlaps the idiotype paratope that disables it from binding the antigen.

**Cytokeratin**

Epithelium is closely packed cells that line and surround organs in the whole body. Stratified epithelia, such as the skin, consist of multiple cell layers whereas simple epithelia, such as the intestinal mucosa, only consist of one cell layer. The epithelia can be differentiated by their profile of a sub-group of intermediate filament proteins, the cytokeratins. The intermediate filaments constitute the cytoskeleton together with the microfilament and the microtubules and are divided into sub-groups where all sub-groups have similar function.

Cytokeratins are rod-like proteins that form hetero-dimeric coiled-coil complexes. The amino acid sequence reveals a heptad repeat (abcdefg)ₙ, where a and d most often are apolar residues typical for a coiled coil protein. There are approximately 20 cytokeratins known, although the number is increasing, and they are divided into two types. Type I is acidic whereas type II is neutral/alkaline and the coiled coil are always built up of one cytokeratin from each type [9].
The most important cytokeratin complex in simple epithelia is the CK8/CK18 complex. This complex is called “primary” keratin complex and are the first cytokeratins expressed in differentiating epithelia followed by other “secondary” cytokeratins. It has also been shown that this “primary” cytokeratin complex is important early in the embryogenesis. It can be detected already in the 4-cell stage in mouse embryos [10].

The function of cytokeratins has been discussed and there are variations in function depending on whether they are found in stratified or simple epithelia. In stratified epithelia the cytokeratins are thought to build up the structure of the cells while the function of cytokeratins in simple epithelia is more complicated. One of the functions that have been proposed is that the cytokeratins are responsible or at least partially responsible for maintaining the cell polarity. It is very important that the cells in simple epithelia have the right cell polarity since they are responsible for secretory and absorptive functions, for example in the intestine [10]. Finally it has been suggested that the cytokeratins may protect the cells from physical stress. One form of physical stress is osmotic stress, which the simple epithelial cells must be highly subjected to. The cytokeratins have the ability to preserve the structure of the cells in the tissue during osmotic swelling or shrinkage [11].

**Cytokeratins in carcinomas**

Epithelial cells are found at locations where they are subjected to high stress and they are normally frequently dividing to replace damaged cells. Epithelial cancers (carcinomas) constitute 90% of all human cancers since the epithelial cells are highly proliferative and vulnerable for carcinogenic transformations [10].

Carcinomas express significant amount of cytokeratins and the necrotic areas in the center of the tumor has a high concentration of the cytokeratins since these proteins have low solubility and will not be released into the circulation [9, 12, 13, 14]. The cytokeratins can be used clinically to differentiate between different types of carcinomas and hyperplasias, and also to indicate metastasis [15, 16].
Protein interactions

Molecular recognition is fundamental in all living organisms and is involved in a wide variety of systems, such as cell signaling, transcription and immunity. This recognition is an important process to understand in molecular biology. A tool in the study of protein interactions is mutagenesis where the side-chain of the amino acid residue of interest is replaced by another residue, often alanine. The effect of the mutation is then studied and the interaction characteristics are compared to the unmutated version of the protein. This technique has shown to be useful in a number of studies [17-20].

In protein-protein complexes, there are usually 10-30 amino acid residues in each protein that contribute to the interaction and the majority of the binding energy is predominantly contributed by three to four residues [21]. An average protein has a residue type distribution of 55% non-polar, 25% polar and 20% charged on the surface and a similar distribution in the interacting interface. No particular residue composition is found except in the case of antibodies, where mostly tyrosines and tryptophanes are found [22].

To form a protein-protein complex, the proteins have to find the proper relative orientation to each other. This includes dehydration and rearrangement of the interface and the formation of short-range interactions. The formation of the complex can be divided into two parts as illustrated in figure 2. First, a weak complex is formed, usually referred to as the encounter complex. Second, the specific docking occurs which yields the high affinity complex. Long-range electrostatic forces drive the formation of the encounter complex. This is followed by a massive dehydration of the interacting interface which contributes favorably to establish other kinds of short-range interactions [18, 19]. There are examples of two different receptors, a complement factor C3d receptor and a steroid receptor, using this two-step approach. These interactions are first established through electrostatic interactions, regulating the association, and are followed by hydrophobic, hydrogen and van der Waals interactions, regulating the dissociation [23, 24].

The electrostatic interactions responsible for the formation of the encounter complex are often net destabilizing in the formation of the tight complex, due to the desolvation. It has been reported that interaction interfaces often is surrounded by a hydrophobic ring that accomplishes the desolvation [25]. To form the protein complex, destabilization of the electrostatic interactions must be overcome and this is accomplished by the hydrophobic effect. The hydrophobic effect is strongly stabilizing and thereby contributes to the complex formation [26].
Figure 2. A schematic illustration of the formation of a tight and stable protein complex. The first step is the long-range electrostatic interactions, which result in the encounter complex. The hydrophobic interactions desolvate the interaction interfaces and short-range interactions, such as hydrogen, van der Waals and hydrophobic interactions, are established and a tight protein complex is formed.

Overall electrostatic interactions are involved in the formation of the encounter complex and are thereby important for the specificity. Furthermore, the electrostatic interactions are also present in the protein complex where they comprise only a part of the complex forming interactions together with the hydrophobic effect, van der Waals interactions and hydrogen bonds [19]. Hydrogen bonds are common in protein interactions, even between backbone atoms [22, 23].

The electrostatic interactions can be improved to have a more favorable effect on the complex formation and the favorable effects of electrostatic interactions can be established in three ways. First, loss of electrostatic interactions to the solvent when binding, desolvation. Second, an intermolecular interaction between functional groups in each protein. Third, an intramolecular interaction between functional groups within the same protein. The third effect, the
intramolecular interaction, is often established between residues just below the interaction interface and residues in the interface [26].

An example where the proteins seem to use optimal electrostatic interactions to accomplish tight complexes is the barstar-barnase complex. Barnase is an extracellular ribonuclease from *Bacillus amyoliquefaciens* and barstar is its natural polypeptide inhibitor. The complex formation of these two proteins is a common system when protein interactions are studied. Barstar and barnase complexes have many salt bridges and hydrogen bonds in the interface and barstar is electrostatically optimized to interact tightly to barnase [17].

The electrostatic interactions are further proven to precede the complex formation in the study of the *Escherichia coli* chaperones GroEL and SecB, which have preference for positively charged substrates, which is coherent with their own negative net charge [20]. The hydrophobic effect stabilizes the chaperones and is preceded by an initial ion interaction.

The theory of an encounter complex preceding the formation of the tight protein complex can of course not be applied on every protein-protein interaction. There are examples of interactions that have a strong and very specific electrostatic complementarity, which form both the encounter and the protein complexes. There are also examples of proteins whose early interactions are driven solely by the hydrophobic effect [27]. However, the encounter complex theory is usable for many protein-protein interactions.

Interacting interfaces are not regions of high flexibility. Despite this, atoms in interacting interfaces very often make a small shift during complex formation. This shift is often a small low energy conformational change [19, 21, 22].

It would seem that a change of the hydrophobicity would be the best if an alteration of the affinity is wanted. If the studied protein is a drug candidate, this might not be the best solution. An increment of the hydrophobicity of a protein can result in protein aggregation, which would be a problem in the drug design.
**Antibody-antigen affinity**

All dissolved proteins, antibodies included, must repel all molecules and cells on a macroscopic level to stay in solution. The specific attraction of an antibody’s paratope to the corresponding epitope must be able to overcome this overall repulsive force to establish binding [28].

When antibodies are used in tumor treatment or tumor targeting, it is desirable to be able to regulate the affinity of the antibody to its antigen and thereby regulate the tumor:non-tumor ratio. An anti-idiotypic antibody can be used to improve the tumor:non-tumor ratio by clearing the circulation from non-tumor binding idiotypic antibody. When the anti-idiotypic antibody has bound to the idiotype, the antibody complex is cleared by the reticuloendothelial system. When an anti-idiotypic is used to clear the idiotype from the circulation, it is even more important to study the interaction and to control it. The affinity between the idiotype and the anti-idiotypic could be so high that the anti-idiotypic empties the tumor on idiotype and makes the immunotherapy impossible. The most desirable scenario would be to have an idiotype where 100% of injected antibody binds to the tumor, but this will not happen in a real situation. The second best is to have an idiotype that as much as possible binds to the tumor and an anti-idiotypic that binds to the rest of the idiotype that is free in the circulation, without affecting the idiotype located to the tumor. This will probably not be the starting conditions for any system. To bring the system closer to the second best choice the interactions have to be studied and certain amino acid residues have to be mutated into other residues to alter the affinity.

If the desired result of a mutation is to increase the affinity, it is most likely that is the electrostatic interactions that have to be altered. The electrostatic interactions are long-range and they have an impact on how long and in what orientation two molecules approach each other. These interactions can make it easier or harder, depending on the electrostatic pattern, for other interactions to take place [29, 30]. There are a number of problems with increasing the affinity. The antibodies are naturally high affinity binders and it is hard to improve something that nature already made good. A mutation could, besides alter the affinity, alter the structure of the antibody and actually destroy the structure of the CDR loops. The effect of a mutation is hard to predict and, finally, the knowledge about antibody antigen interactions is not very well understood.

It is much easier to decrease the affinity than to increase it. Usually there are 12-20 residues that form the interaction surface between the antibody and its antigen. These residues contribute with several van der Waals interactions and some hydrogen bonds and electrostatic interactions. These residues are very
important for the interaction and will therefore almost always result in a decreased affinity if mutated [31].

Tools for studying antibody interactions

Site directed mutagenesis is a powerful tool in the study of antibody-antigen interactions, the so-called functional epitope/paratope can be identified. The effect of the mutation can be studied using different techniques, such as ELISA and SPR. Different mutagenic strategies can be applied. One common way in mutagenic studies is the “alanine walking” where every residue, one by one, is mutated to alanine. It can be wise to precede the mutagenic procedure with some kind of guidance of which amino acid residues that will have the greatest effect on the interaction when mutated. One approach is to make mutations on the basis of statistical studies such as MacCallum et al. [32] or to build a model and mutate on what seems to be important residues. Groups of residues can also be chemically modified, for example acidic residues, and the effects can then be studied. The goal of chemical modification is to modify specific residues, giving an altered protein without affecting other residues or causing a conformational change [33].

The mutational system reveals the functional paratope rather than the classical structural paratope. The functional paratope is much smaller than the structural paratope since a large area of the antibody can be adjacent to the antigen, the structural paratope, but only a few residues spread out over that area are actually involved in the interaction, the functional paratope [34].

Antibody modeling

Two well known methods to identify protein structures are X-ray crystallography and nuclear magnetic resonance (NMR). These methods are generally slow and difficult, whereas the cheaper and faster computer modeling can be a good alternative.

The most common way to make a model is by homology modeling. In this procedure the protein of interest is modeled against a template protein structure based on residue sequence identity and similarity. The success in the search for appropriate templates is crucial since the reliability of the model depends on the template. For proteins that share 50% of its sequence with the template, the
model deviate approximately 2 Å from the true structure [35] and less sequence identity gives higher deviation.

There are various softwares available that can be used in computer modeling but an alternative is the web based modeling servers such as Swiss-Model which is a homology modeling server. This service also includes a search engine to search the Protein Data Bank (PDB) [36] for suitable protein templates [37].

Modeling of antibodies are somewhat easier than modeling of the average protein. There are many X-ray and NMR structures of antibodies in PDB and the antibodies are well defined molecules. The CDR classification with canonical classes and clusters are making the models even more precise. The backbone deviation of the antibody model from the true structure is typically 1.0 Å in the non-CDR (frame) region, assuming that the template is correct and determined with high resolution.

As discussed earlier, the CDRs can be designated a class and cluster, which can be very useful in the antibody modeling. If the CDRs are modeled against CDRs from the same classes and clusters, the model will be even more precise than described above. As good as 0.2 Å deviation has been accomplished [38]. The problem in CDR modeling is to find CDRs from the same class and cluster that is structurally determined with good resolution. This is of course a problem when it comes to CDR 3 in the heavy chain because of its diversity in length and sequence.

The protein model must be evaluated in some way to establish whether the model is reliable enough for the specific purposes one might have. The conformation of the amino acid residues and the bond length must be controlled. The model can be compared to known structures of related proteins. There are programs, like for example WhatCheck, which can make this evaluation [39]. The programs control the chemistry in the model and not the biological function of the modeled protein. A protein model can be constructed in a way that would give no biological function, but is completely correct concerning the chemical constraints.

When the model is done, it can be presented in a number of ways depending on what point the scientist is trying to prove with the model. It can be presented with the classic sticks and balls or space filling. It can also be presented with different kinds of surfaces, van der Waals surface, molecular surface or solvent accessible surface. The model can then be colored to present electrostatic potential, hydrophobicity, and polarity amongst other things.
A warning must be raised when handling models. It is easy to consider the models to be the real and true structures of the protein. This is not the case. The models must be used as models for guidance in what direction further actions might be taken. It is also important to remember that all structures, modeled, x-ray and NMR, are frozen moments in time. The real protein structure is not static but highly flexible.

Aims of this study

The aims of this study were to first, synthesize and produce a functional scFv of the mouse monoclonal anti-cytokeratin 8 antibody TS1. Second, functionally map residues that are important for the interaction with the antigen, cytokeratin 8, and to the anti-idiotype, αTS1.
Results and discussion

This study focuses on the interaction of the anti-cytokeratin 8 scFv TS1-218 with its antigen and its anti-idiotype αTS1 by studying the effect of chemical modifications of CK8 and the IgG’s TS1 and αTS1 and the effect of site-directed mutagenesis of the scFv TS1-218 in ELISA. It also presents how the scFv were constructed from the monoclonal mouse IgG.

Two scFvs were made, but only one was functional. First, the scFv TS1 was constructed in the V_{H}-linker-V_{L} order with the common (Gly4Ser)₃ linker [8]. However, when this scFv was expressed, it was not functional. Second, the scFv TS1-218 was constructed in the V_{L}-linker-V_{H} order using the 218 linker (GSTSGKGPGSGEGSTKG) described by Filpula et al. [40]. The scFv TS1-218 was fully functional with affinity to both its antigen CK8 and its anti-idiotype αTS1. The amino acid sequence of TS1 V_{L} and V_{H} numbered according to Chothia and Lesk 1987 [2, 41] is displayed in table 1, together with MacCallum et al. statistics [32].

Table 1. Amino acid sequence of the TS1 V_{L} and V_{H} numbered according to Chothia and Lesk 1987 [2, 41]. Bold residues are included in the central interacting interface in TS1 (paper II). Underlined residues show the residues that were mutated (paper II). Underlined numbers show residue positions where at least 50% of the CDRs in the MacCallum et al. statistics [32] interact with the antigen.
The antibody model made in paper I and further developed in paper II, displays a mostly polar surface which is broken by smaller hydrophobic patches (figure 3). Some of the hydrophobic residues constitute a ring surrounding the area believed to be the interacting interface. The surface arising from the CDRs displays four negative charges and one positive charge. Furthermore, nine tyrosines and one tryptophan are exposed in the CDR surface.

In the model, a deep pocket is displayed just north of the center of the CDR surface (figure 3a). To the south and east of this pocket is the CDR 3 from the heavy chain located. This CDR is generally considered to play a major part in the antigen interaction.

**Figure 3. a:** Surface homology model of the mab TS1 variable domains viewed from the interaction site. A line is surrounding the defined central interacting interface (paper II). Y97H3 is important for the interaction to CK 8 and W100bH3 is important for the interaction to αTS1. **b:** The epitope peptide of CK8 (QRGELAIKDANAKLSELEAALQRAKQ) presented as an opened α-helix. The residues most important in the CK8 – TS1 interaction are indicated with a circle. Both images have the residues coloured by type. Hydrophobic – grey, polar – yellow, acidic – red, basic – blue.
The antibody model made in this study must be considered to be relatively accurate. The variable domains of TS1 have approximately 55% sequence identity with the template sequence and are thereby not likely to deviate more than 2 Å from the true structure according to Schwede [35]. The class and cluster could be identified correctly for LCDR 2 and 3 and cluster but not class could be identified for LCDR1 and HCDR 1 and 2. This decreases the deviation further. LCDR 2 and 3 with fully identified class and cluster are probably modeled with very high accuracy whereas their relative positions are less accurate.

When using models, an update will have to be performed from time to time because more and more structural data becomes available over time. The new structures can result in even better models especially if a higher sequence identity between the modeled protein and the templates can be established. When searching PDB for suitable model templates for TS1 V\textsubscript{H} and V\textsubscript{L}, there are now (September 2004) structures available with higher sequence identity than when the model first was made in late 1999. The sequence identity is now 88% whereas it was 55% in 1999. This increase in sequence identity would bring the deviation closer to 1 Å than 2 Å, as previously predicted. The CDRs would, as in paper I, be modeled against templates based on class and cluster which would decrease the deviation further for these parts of the antibody. There can also be antibody structures available that have CDRs representing a class or cluster whose structure was not determined earlier. When classifying the CDRs of TS1 (in September 2004) in Martin's bioinformatics site [3], the result is basically the same as in paper I. The only possible change is for LCDR1 where the class still is unidentified but the cluster could possibly be changed from 15B to 15A. This change is not certain since LCDR1 of TS1 does not fit exactly into any of these two clusters.

Chemical modifications of TS1 IgG performed in paper I were used to determine which residues or groups of residues that are important for the interaction. Five reagents were used to perform the chemical modifications. EDC mainly modifies aspartic/glutamic acid and tyrosine, DEPC mainly modifies histidine and primary amino groups, pHPG modifies essentially arginine. FDNB modifies mainly tyrosine, lysine, histidine cysteine and primary amino groups. The acetic anhydride modifies essentially lysine and primary amino groups. TS1 IgG was adsorbed to microtiter plates and then subjected to one of the five reagents. The modifications were made with and without protection of CK8 and αTS1 to distinguish between modifications on functionally important residues and general denaturation. It was suggested that the important residues in TS1 were tyrosine and aspartic and/or glutamic acids for the interactions with both CK8 and αTS1. These results were confirmed, in
paper II, since several of the mutated tyrosines and acidic residues decreased the affinity.

The modifications indicate that there are acidic residues that are involved only in the CK8 interaction and only in the αTS1 interaction respectively. The interaction sites on TS1 with CK8 and αTS1 are likely overlapping but not identical since the modifications of TS1 gave a similar result for some modifications and different result for other modifications, when binding to CK8 and αTS1 respectively.

The N-terminal amino acid residues are important in two different aspects. First the N-terminal amino group, the chemical modifications of TS1 IgG suggested that at least one of the N-terminal amino groups was important for the interaction to both CK8 and αTS1. Since the first scFv construct did not function when the light chain nitrogen in the N-terminal amino group was occupied in the peptide bond to the linker it was concluded that it is the N-terminal amino group in the light chain that is important for the function of the scFv.

The other aspect concerning the N-terminal amino acid residues involves the side-chains of the residues. When the N-terminal acidic residues (VH1E1 and VH1D1) were mutated to amides the affinity increased significantly as suggested by the chemical modifications. Although, in addition to these results, the most important finding is to not have an acidic residue N-terminally because the mutations of these residues to alanine or respective amide gave a similar increase in the affinity. Furthermore, the mutation of VH1Y58 to alanine resulted in increased affinity to CK8. All three of these mutated residues are likely to be located outside the interaction interface since it has been shown that residues able to increase the affinity when mutated do not take part in the actual interaction, but improve the preceding steps [29, 30]. The location of these residues on the surface are shown in figure 3.

On the basis of the chemical modification and the gel permeation chromatography experiment described in paper I, together with the MacCallum et al. statistics [32], it was concluded that the residue VH1H34 could be important for the interaction of TS1 with αTS1. For the site-directed mutagenesis experiments with VH1H34 mutants, the interaction with both CK8 and αTS1 was affected giving contradicting results. This could be explained by the differences in the two systems. In the chemical modification, reagents specific for different functional groups of a residue modify that residue, whereas in the mutation procedure the side chain is substituted to a single methyl group. The accessibility of the residues must also be considered. The mutation is confirmed by sequencing but the chemical modification is not really confirmed and if a
side chain is buried and not accessible for the chemical compound the modification of that side chain will not occur. According to the model, created in paper I and further developed regarding accessibility in paper II, it is obvious that this residue (VlH34) is not accessible from the surface which would imply that the chemical modification did not occur for this residue (figure 2b paper II). The gel permeation chromatography of TS1-αTS1 complexes formed at two different pH suggests that a mutation of the VlH34 to a lysine or arginine might improve binding to CK8. After the study of the accessibility these mutations no longer seem to be of interest.

In TS1, residues important for interaction with both CK8 and αTS1 were identified and included in a central interacting interface, displayed in figure 3. Some of these residues are likely hotspots which are defined as a residue contributing energetically >2kcal/mol to the interaction. The central interacting interface is made up of residues that are located close to each other and when mutated, the affinity will be abolished or substantially decreased. Interaction sites are most often made up of by tyrosines, arginines, tryptophanes and in some cases serines [25, 42, 43]. In the central interacting interface of TS1 there are mainly these types of residues but also a glutamic acid and a histidine. The histidine (VlH34) is located centrally in the interacting interface in a deep pocket and is concluded not to participate in the interaction but to have a structurally important function. It has been suggested that when CDR 3 in the heavy chain is of medium size, as it is in TS1, residues in the positions Vl34 and VH100a interact to accomplish the structure required for the antigen interaction [44]. This in combination with the buried location of the VlH34 in the model of TS1 indicates that this residue is more likely to have a structural important function than being functional for the interaction.

It is also shown that the residues constituting the interaction interface are surrounded by a hydrophobic ring. This ring will desolvate the interaction interface and thereby prevents the water molecules in the solvent to interact with the residues in the interface. The desolvation of interaction sites is a necessity for energetic interactions and is supported by both experimental data and computer models of other proteins than antibodies [25, 26, 45].

In most antibodies the LCDR 2 is not involved in the interaction but contribute to the structure required for antigen interaction. Despite this, two amino acid residues in LCDR 2 of TS1 are considered to be important for the antigen binding. Furthermore, these two residues are surrounded by the hydrophobic ring and are located in the area defined as the central interaction interface. In addition, it has been reported of LCDR2 involvement in the interaction when the antigen is large [32] as in the case of this study.
When studying the MacCallum et al. statistics [32] in table 1, it might be concluded that more residues should be analyzed by mutagenesis. This might be a good idea, but the model of TS1 suggests that several of the statistically important residues are completely buried and these residues are often flanked by almost completely buried residues.

When studying the surface of the model of TS1 in figure 3 together with the epitope of CK8 a question might be raised. They both show a similar pattern of negative charges that logically should repel each other and interaction between the molecules would not occur. There is however, no doubt that TS1 bind to CK8. There are reports of a stabilizing effect in coiled coil proteins using salt bridges between charged side-chains [46, 47]. In paper II, the web-interface of the software What If [48] reported three possible salt bridges (E343-K347, D348-K352, E355-R362) which could “neutralize” at least two of the three negative charges in the CK8 epitope and have a positive effect on the interaction of TS1 with CK8.

**Future perspective**

In the future, the scFv TS1-218 is to be tested in vivo for tumor therapy. To use a scFv in tumor treatment it has to perform some cytotoxic effect. In our case, an effector molecule has to be coupled to the scFv to accomplish this effect. We have chosen the β-glucuronidase enzyme from E.coli to be coupled to the scFv TS1-218. This enzyme can, in the tumor, convert a non-cytotoxic drug to a cytotoxic drug and thereby kill the tumor cells.

It is in this study suggested that the mutation of V15Y58 to alanine would improve the affinity to CK8 and decrease the affinity to αTS1. This can be used to improve the tumor:non-tumor ratio of scFv TS1-218 in vivo. The suggested mutation will direct more of the scFv into the tumor compared to the unmutated form. The affinity to αTS1 is relatively high so the decrease in that interaction will probably not be a problem.

Furthermore, it would be interesting to perform kinetic studies on some of the mutated versions of scFv TS1-218. The association and dissociation rates and not only the affinity are important if the antibody is to be used in tumor therapy. To obtain an even distribution of antibody in a tumor, the association rate can not be too high and the dissociation rate can not be too low for the interaction. The kinetic study would also make it possible to calculate energy contribution in the interaction from specific residues, thereby making it possible to identify hotspots.
Acknowledgments

It would not have been possible for me to do this work without a number of persons who really deserves my gratitude.

Associate Professor Birgitta Sundström for inviting me to work in her project. If you count the years it seems like a long time ago, but it certainly doesn’t feel that way. Thank you for sharing all your skills and experience. Protein purification is not that easy really!

Professor Thomas Nilsson and Associate Professor Leif Jönsson for making valuable comments on my work. Thank you!

Professor Torgny Stigbrand at Umeå University for your knowledge, hospitality and of course for the supply of the Biacore. Many thanks to you and to the whole staff in Umeå for making it a pleasant visit despite the “cubby-hole”.

Thank you Ann for sharing of your skills and ears when the antibodies didn’t behave like we wanted them to.

Rozbeh, thanks for sharing the mutant work and for always having a smile on your face. I hope you will join us again!

To the colleagues at the Division for Chemistry and of course, especially the Department of Biomedical Science, for making the University a nice place to work. Thank you all!

Many thanks to my close friends and family for making me live and breath something else than protein structures and antibody modeling. Thank you!

Last but not least, Sarah for making my life much more enjoyable. I love you!
References


42. Padlan E.A. (1990) On the nature of antibody combining sites: unusual structural features that may confer on these sites an enhanced capacity for binding ligands. Proteins. 7:112-24.


Single-chain antibody construction and functional mapping of the monoclonal antibody TS1

The aims of this study are to synthesize and produce a single-chain antibody (scFv) of the anti-cytokeratin 8 monoclonal IgG antibody TS1 and to functionally map amino acid residues important for the interaction with its antigen and the anti-idiotypic antibody TS1. The TS1 antibody has been shown to be effective in binding cytokeratin 8 (CK8) expressed in tumors in vivo and is proposed to be useful in immunotargeting and/or immunotherapy. The anti-idiotypic antibody TS1 can be used to regulate the tumor:non-tumor ratio. Mutagenesis of certain amino acid residues can be used to alter the affinity to improve the tumor:non-tumor ratio further.

In the present study, the TS1 IgG was chemically modified to specify groups of residues important for interaction with both CK8 and TS1. If important residues were found in the CDRs, they were mutated in the TS1 scFv construct and the effect was studied using ELISA.

The main conclusions drawn from this study are that the important amino acid residues in TS1 for the interaction with both CK8 and TS1 are mainly tyrosines, charged residues and a tryptophan. A central interacting interface was identified with the somewhat unusual participation of residues in the CDR 2 of the light chain. Mutations which resulted in increased affinity to both CK8 and TS1 were also identified.